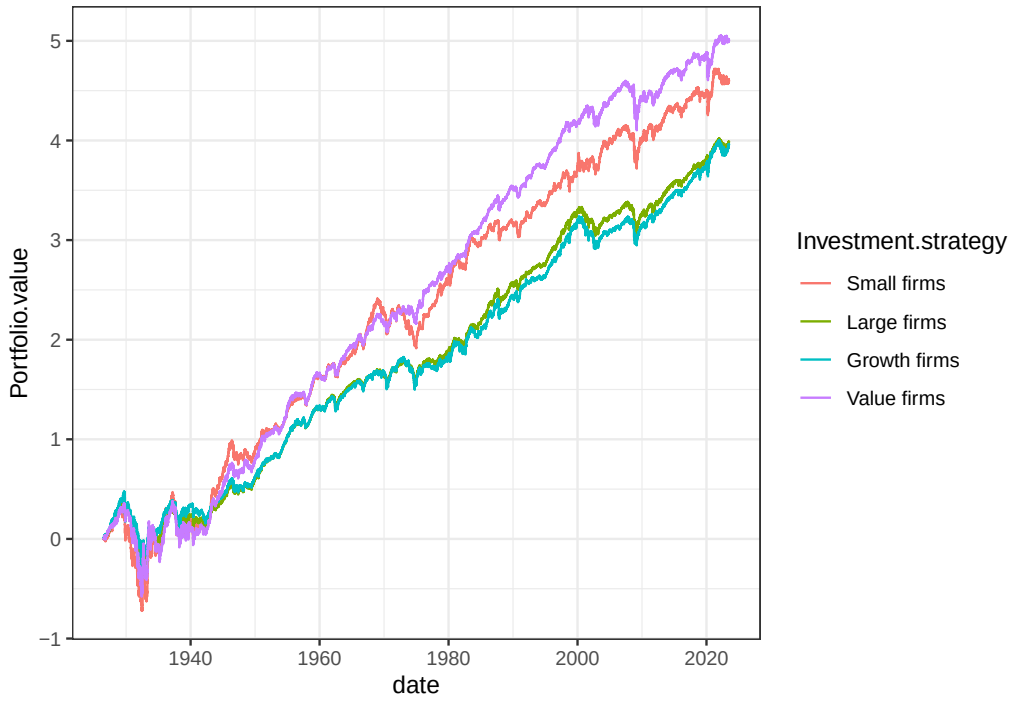
Financial Econometrics

Simple Regression

Dmitrii Pugachev

Two of the most well-known facts in Finance are: (1) small stocks earn higher returns than big stocks, and (2) stocks with high book-to-market (BM) ratio [value stocks] earn higher returns than stocks with low book-to-market ratio [growth stocks]. The plot below illustrates these two facts by showing the value of four portfolios from 1 July 1926 till 30 June 2023. The red (green, blue, purple) line shows profits from a value-weighted portfolio of 30% smallest (largest, highest BM, lowest BM) firms. Each portfolio starts with investing $1 in the corresponding strategy. Portfolios are re-balanced every year on June 30. Note that the plot uses log 10 scales for the y-axis. By 30 June 2023, portfolio of small (value) stocks outperforms portfolio of large (growth) stocks on $30 594 ($94 125).



In this lab session, you will use data on S&P500 firms from 2010 to 2015 to revisit the aforementioned findings in this particular data set. In particular, using the fiscal year 2014, you want to estimate the following two regression equations:

$$Return_{Next\ fiscal\ year} = \alpha + \beta\ log(BM) + u$$

$$Return_{Next\ fiscal\ year} = \alpha + \beta\ log(MKTV_{End\ of\ current\ fiscal\ year}) + u$$

This exercise takes you through the standard steps in econometric analysis: load data, construct variables,

limit the sample, and analyse.

You are provided with a csv file `Data_lab_simple_regression.csv`. The file includes the following variables:

- lpermno: firm identifier
- datadate: date when the data was supplied to authorities
- fyear: fiscal year
- BE: book value of equity
- MKTV.june.of.current.fiscal.year, MKTV.end.of.current.fiscal.year, MKTV.end.of.subsequent.fiscal.year: market value recorded at different times
- FF.49: industry classification following Fama and French's 49 industries
- busdesc: description of the business of the firm

Your analysis proceeds in the following steps:

- Task 1: Load the data into R.
  - Investigate the structure of the data [use the function `str()`]
  - Summarize the data [use the function `summary()`]
  - [optional] Change the date format [use `as.Date()`]
- Task 2: Construct new variables
  - Book to market is defined as book value of equity divided by market value of equity as of June of the current fiscal year
  - Return during the next fiscal year is defined as the percentage change of market capitalization between the end of subsequent fiscal year and the end of current fiscal year
- Task 3: Prepare the required variables for analysis
  - Find summary for log(MKTV.end.of.current.fiscal.year) and construct its histogram [use `hist()`]. Verify that the obtained picture resembles a normal distribution.
  - Find summary for book to market, construct a histogram. What are the issues with its lowest and largest values? Replace values below the 2.5% percentile with the value of the 2.5% percentile, and then replace values above the 97.5% percentile with the value of the 97.5% percentile [use `quantile()` function, check option "prob"]. This operation is called "winsorization" at 2.5%. Construct histogram for corrected book-to-market and for its logarithm. Verify that log(book to market) looks good.
  - Find summary for returns, construct a histogram. Does the picture look good? Apply winsorization at 2.5% and then construct the histogram. Is it better now?
- Task 4: Estimate the regression equations as outlined above, using only the fiscal year 2014. Note that you can use the logarithm function `log()` within the regression function `lm()`.
- Task 5: Having estimated both models, combine them in a output table with the package `stargazer`. Check `?stargazer()` (section "Examples") to learn how to combine several models in one table.
- Task 6: Complete following sentences based on your findings,
  - An increase in log(BM) on 1 leads to a ... [increase/decrease] of returns by ...% at the ...% significance level.
  - An increase in log(MKTV) on 1 leads to a ... [increase/decrease] of returns by ...% at the ...% significance level.

  Are your findings in line with the plot from above (value stocks outperform growth stocks, small stocks outperform large stocks)? If not, why could this happen? Is this test stronger than the test from lab session for week 2?
- Task 7: (Optional) You want to know, if the estimated coefficient $\hat{\beta}$ for the relationship between BM and future return depends on the year the regression is estimated for. Re-estimate equation 1 for each fiscal year from 2010 to 2014 and report in a stargazer table. Does the relationship between BM and future return seem to be robust? Repeat the task for the second regression.